

Aspektusszintű annotáció és szentimentet módosító elemek egy magyar nyelvű szentimentkorpuszban

Szabó Martina Katalin^{1,2}, Vincze Veronika³, Hangya Viktor⁴

¹Precognox Informatikai kft.

²Szegedi Tudományegyetem, Orosz Filológiai Tanszék
mszabo@precognox.com; szabo.martina@lit.u-szeged.hu

³MTA-SZTE Mesterséges Intelligencia Kutatócsoport
vinczev@inf.u-szeged.hu

⁴Szegedi Tudományegyetem, Informatikai Tanszékcsoport
hangyav@inf.u-szeged.hu

Kivonat: A dolgozat egy magyar nyelvű kézzel annotált szentimentkorpusz annotálásának tapasztalatait, valamint hasznosításának első eredményeit mutatja be. A dolgozat célja kettős: Egyrészt, bemutatjuk mindazokat a módosításokat, amelyeket a korpusz második annotálási fázisában alkalmaztunk az első annotálási fázis megoldásaihoz képest, a későbbi hasznosíthatóság javítása érdekében. Másrészt, megmutatjuk a korpusz első vizsgálati és felhasználási eredményeit. A vizsgálat keretében elsősorban a különböző annotált elemek fragmentumbeli, valamint szófajok közötti megoszlását tártuk fel. A dolgozatban beszámolunk a korpusz kvantitatív, illetve kvalitatív sajátosságainak néhány figyelemre méltó összefüggéséről. Mindemellett az annotációra támaszkodva bizonyos elemcsoportokból lexikonokat generáltunk, amelyek reményeink szerint támogatni tudják majd az automatikus szentimentelemzés eredményességét. A dolgozat a generált lexikonok alapvető adatairól is számot ad.

1. Bevezetés

A dolgozatban a magyar nyelvű kézzel annotált szentimentkorpuszunk annotálásának tapasztalatait, valamint hasznosításának első eredményeit mutatjuk be.

Egy korábbi dolgozatunkban beszámoltunk egy magyar nyelvű szentimentkorpuszról, amelyet termékvélemény-szövegekből hoztunk létre kutatási és fejlesztési céllal [1]. A korpusz szöveganyagát a [<http://divany.hu/>] honlap termékvéleményeiből állítottuk össze. A honlap készítői időközönként bizonyos termékcsoportokat tesztelnek, s közzéteszik a tesztelők véleményét. (1) alatt közlünk két példát a nyers korpuszból.

- (1) a. Elképesztően kellemetlen íze van, előfordult, hogy valaki nem is ismerte fel benne a csokoládéízt. Legtöbbször az hangzott el, hogy mellékíze van, édes és bővli termék. Sajnáljuk azokat a gyerekeket, akik ezt kapják. Igaz, a csomagolása elég kecsesgető, de ez ne tévesszen meg senkit!
[<http://divany.hu/2013/karacsony/adventinaptarteszt/>]

b. Nem maradt kellemetlen, zsíros érzés a bőrömön használat után. Az illata nagyon kellemes, nem kell nagy mennyiség belőle és úgy érzem, jól tisztít. Nem szárítja a bőröm, így használata után nem is kell hidratálót felkennem, üde, jó érzés. Ráadásul nagyon olcsó, az 50 milliliteres mini kiszerelés csak 199 forint!

[http://divany.hu/tejbenvajban/2014/08/28/arclemoso_teszt_tej_vagy_gel]

A korpusz második annotálási fázisában átdolgoztuk az annotálási elveket, majd a korpusz teljes szöveganyagát feldolgoztuk kézzel.

A korpusz hiánypótló magyar nyelvű kutatási és fejlesztési eszköz. Tudomásunk szerint ez az egyetlen olyan magyar nyelvű szentimentkorpusz, amely fragmentum- és aspektusszintű annotációval rendelkezik, emellett az annotáció keretében számos, a szentimentelemzés szempontjából releváns nyelvi elem is önálló taget kapott (az annotációról, valamint annak elméleti alapvetéseiről részletesebben l. később, 2).

Az elkészült adatbázis összesen 154 véleményyszöveget, 17 059 mondatot és 251 202 tokenet (központozással) tartalmaz.

2. Az annotálási rendszer felülvizsgálata

2.1. Annotációs alapelvek

A korpusz első annotálási fázisában a teljes értékelő kifejezést (másképpen: szentimentfragmentumot), azon belül a pozitív és a negatív polaritású szentimentkifejezéseket, azok targetjeit, valamint esetleges siftereit jelöltük be a korpuszban [1].

A szentimentelemzésben a feldolgozási egységek terjedelme szerint alapvetően a következő három szintet különböztetjük meg: a dokumentum-, a mondat-, valamint az entitás- és aspektusszintű annotációt [2]. A dokumentumszintű elemzés (*Document-level Sentiment Classification*) célja, hogy a teljes szöveg viszonylatában megadja a benne megfogalmazott értékelés polaritását, tehát azt, hogy az adott szöveg összességében pozitív vagy negatív értékelést fejez-e ki. A mondat szintű elemzés (*Sentence-level Sentiment Classification*) a szöveg egyes mondatainak vonatkozásában kívánja meghatározni azok polaritását. A mondat szintű elemzés egy altípusának tekinthetjük az ún. tagmondat-szintű elemzést (*Clause-level Sentiment Classification*), amely az értékelés polaritását az egyes tagmondatok vonatkozásában határozza meg. Megállapítható, hogy bár a mondat-, illetve tagmondat szintű osztályozás kétségtelenül árnyaltabb feldolgozást tesz lehetővé a dokumentumszintű osztályozásnál, igazán hatékony szentimentelemzést egyik megoldással sem érhetünk el. Tudniillik, egyik fentebbi eljárásmóddal sem tárhatjuk fel azt, hogy a szövegben megfogalmazott értékelés pontosan mire, azaz mely targetre irányul, emellett további problémákat eredményez, ha az adott, egyetlen egységként kezelt szövegrész (dokumentum vagy mondat) több targetet és/vagy több szentimentet is tartalmaz.

A fentebb elmondottakkal összefüggésben, a leghatékonyabb szentimentelemzési megoldást az ún. entitás- és aspektusszintű feldolgozás (*Entity and Aspect-level Sentiment Classification*) kínálja. Ez a megoldás ahelyett, hogy a szöveg valamely szerkezeti egységét (dokumentum, bekezdés, mondat vagy tagmondat) venné az elemzés alapjául, kifejezetten magukra az értékelésekre koncentrál, és a feldolgozás alap-

egységét egy target, valamint az annak vonatkozásában kifejezett szentiment kapcsolatában határozza meg [2].

A bemutatott megközelítési módok hatékonyságát összevetve úgy döntöttünk, hogy korpuszunk annotálásában entitás- és aspektusszintű feldolgozást alkalmazunk. Ez a gyakorlatban azt jelentette, hogy egy egységként, azaz szentimentfragmentumként annotáltunk minden olyan szövegrészt, amely egyetlen meghatározott polaritású értékelést fejezett ki egyetlen (esetleg több, egymással mellérendelő szintaktikai viszonyban álló) target vonatkozásában. E megoldás következtében a fragmentumok az esetek túlnyomó többségében mondatnyi, vagy a mondatnál kisebb egységek, azonban ritkán az is előfordul, hogy egy fragmentum átível a mondat határán.

A szentimentfragmentum annotálását követően, azon belül több, a szentimentelemzés szempontjából releváns nyelvi elemet eltérő taggel láttunk el. Szentimentkifejezésként kezeltük azokat az egy szóból álló, vagy állandósult többszavas szókapcsolatokat, amelyek lexikai szinten értékítéletet hordoznak valamely target vonatkozásában [1]. Azokat a nyelvi elemeket, amelyek valamilyen módon hatást gyakorolnak a szövegekben megfogalmazott értékelő tartalmakra, az angol nyelvű terminológia alapján szentimentshiftereknek neveztük, és külön taggel láttuk el a korpuszban [2]. Ily módon a negáló, az irreáló, valamint a növelő és a csökkentő értelmű intenzifikáló elemek önálló jelölést kaptak az annotációban.

2.2. A második annotálási fázisban alkalmazott módosítások

A második annotálási szakaszban bizonyos mértékben eltérő annotálási megoldásokat alkalmaztunk a korábbi feldolgozási rendszerhez képest. Ennek oka az volt, hogy a megelőző munkafázis, valamint az első korpuszvizsgálatok tapasztalatai alapján úgy véltük, a targetek kezelésében kidolgozottabb, még részletesebb elemzést lehetővé tevő annotációra van szükség.

Amint azt a megelőző részben ismertettük (l. fentebb, 2.1), a korpusz első feldolgozási fázisában, a fragmentumszintű feldolgozásnak köszönhetően már a targetekkel összefüggésben kezeltük a szentimenteket. Ugyanakkor az entitás- és aspektusszintű annotációt nem tudtuk maradéktalanul megvalósítani, tekintettel arra, hogy az entitásokat, valamint azok egyes aspektusait nem különböztettük meg az annotáció szintjén egymástól. Az újabb annotációs alapelv szerint a korpuszban előforduló entitásokat, valamint azok különböző aspektusait eltérő annotációs taggel (target 1-20) láttuk el, következetesen alkalmazva azokat egy adott dokumentumon belül. A <target 1> címkével rendre az adott termék, azaz maga az entitás szövegbeli előfordulását jelöltük, míg a többi target-címkével annak különböző aspektusait annotáltuk. A címszerűen előforduló terméknevek jelölésére, akárcsak a korpusz első annotálási fázisában, a topic-címkét alkalmaztuk.

A bemutatott feldolgozási megoldás kettős haszonnal bír. Egyrészt, a módosítás eredményeképpen olyan annotáció jött létre, amely lehetőséget ad az entitás-aspektus-összefüggések gépi feltárására, segítve ezzel egy hatékonyabb automatikus szentimentelemző rendszer kidolgozását. Az entitások és az aspektusok megkülönböztetése a szentimentelemzés szempontjából kardinális probléma. Nagy jelentőséggel bír ugyanis, hogy egy értékelés az adott entitáshoz, vagy csupán annak valamely aspektu-

sához kapcsolódik-e, illetve, hogy a szövegben megjelenő aspektusok közül pontosan melyiket minősíti [2]. Tekintsük az alábbi példákat!

- (2) a. Bár a *kiszolgálás* nem olyan jó, imádom ezt az *éttermet*.
 b. A *kiszolgálás* nem olyan jó, a *berendezés* viszont barátságos.

Mindkét példa két ellentétes polaritású értékelést fogalmaz meg. Azonban a (2a) alatti példában a pozitív szentimentérték jelentősebb a negatív szentimentértéknél, tekintve, hogy az előbbi a teljes entitáshoz, az utóbbi pedig csupán annak egy aspektusához kapcsolódik. Ugyanakkor a (2b) alatti példában mind a pozitív, mind a negatív értékelés egy entitás egy-egy aspektusára vonatkozik.

A fentebb ismertetett hasznon túl, az új annotációs megoldás még a targeteket érintő koreferencia-viszonyok kezelését is biztosítja, hiszen egy dokumentumon belül egy adott entitást, valamint egy adott aspektust (és az esetlegesen azokat helyettesítő névmásokat) rendre ugyanazzal a target-taggel láttunk el.

3. A korpusz hasznosításának első eredményei

3.1. Vizsgálati eredmények

A korpusz első vizsgálatának keretében a különböző annotált elemek pozitív és negatív fragmentum-beli, valamint szófajok közötti megoszlását tártuk fel. A korpusz vizsgálatára azért volt szükség, hogy pontos képet kaphassunk a szentimentkifejezések használati sajátságairól a magyar nyelvű szövegek vonatkozásában.

Az alábbi táblázat összefoglalja a korpusz alapvető statisztikai adatait:

1. táblázat. Az annotáció alapvető statisztikai adatai

| annotált elem | összes előfordulás | pozitív fragmentumban | negatív fragmentumban |
|------------------|--------------------|-----------------------|-----------------------|
| PosSentiment | 6693 | | |
| NegSentiment | 8053 | | |
| SentiWordPos | 7554 | 5800 | 1754 |
| SentiWordNeg | 7698 | 1272 | 6426 |
| Topic | 1365 | | |
| Target | 7827 | 3731 | 4096 |
| Negation | 3342 | 1360 | 1982 |
| IntensifierPlus | 4973 | 2405 | 2568 |
| IntensifierMinus | 1080 | 301 | 779 |
| Irreal | 991 | 296 | 695 |
| OtherShifter | 1134 | 642 | 492 |
| ÖSSZES: | 50747 | 15807 | 18792 |

A fentebbi adatok alapján a következő legfontosabb megállapításokat tehetjük: A pozitív (SentiWordPos) és a negatív szentimentkifejezések (SentiWordNeg) majd-hogynem egyenlő arányban képviseltetik magukat a korpuszban, ugyanakkor ez nem eredményezi azt, hogy a pozitív (PosSentiment) és a negatív fragmentumok (NegSentiment) is hasonló megoszlási arányt mutatnának. A negatív fragmentumok ugyanis többségben vannak a pozitív fragmentumokkal szemben. Az eredmény a shifterek szentimentelemzés-beni szerepére mutat rá, amelyek képesek a szentimentkifejezések lexikai szintű polaritásának akár teljes mértékű megváltoztatására.

A fentebbi eredményekből következően, a pozitív és a negatív szentimentkifejezések megoszlása a velük azonos polaritású fragmentumokban nem egyforma. Amíg ugyanis a negatív szentimentkifejezések 83,47%-a azonos polaritású fragmentumban szerepel, azaz csupán 16,52%-uk található pozitív fragmentumban, addig a pozitív szentimentkifejezések magasabb megoszlási aránnyal (23,21%) fordulnak elő a velük ellentétes polaritású fragmentumokban. A kapott eredmények alapján, egy negatív értékelést pozitív szentimentkifejezéssel (pl. *nem jó*) gyakrabban fogalmazunk meg, mint egy pozitív értékelést negatív szentimentkifejezéssel (pl. *nem rossz*). A tapasztalatok illeszkednek az ún. Pollyanna-hipotézishez, amely nyelvi univerzáléként tételezi a pozitív töltetű kifejezések magasabb használati arányát a negatív töltetű nyelvi elemekkel szemben [3].

Kíváncsiak voltunk, vajon mutatkozik-e valamilyen megoszlási eltérés a növelő és csökkentő értelmű intenzifikáló elemeket illetően a pozitív és a negatív fragmentumok között. Megvizsgálva a korpuszban annotált intenzifikáló elemeket megállapítottuk, hogy a növelő értelmű intenzifikáló elemek arányaikban közel azonos gyakorisággal fordulnak elő a pozitív (6693:2706) és a negatív (8053:3347) polaritású fragmentumokban (pl. *nagyon jó*; *nagyon rossz*). Ugyanakkor, a csökkentő értelmű intenzifikálók előfordulási gyakorisága nem azonos a két fragmentumtípusban: a negatív fragmentumokban (8053:779) jóval gyakrabban fordulnak elő a pozitív fragmentumoknál (6693:301). A jelenségnek valószínűleg az az oka, hogy amíg a növelő értelmű intenzifikáló elemek nem képesek a szentimentérték megváltoztatására, addig a csökkentő értelműek igen, és jelenlétük nem azonos szemantikai változást eredményez a pozitív és a negatív polaritású szentimentkifejezések mellett. Amíg ugyanis a pozitív polaritású szentimentkifejezések módosítójaként rendre negatív értékítéletet eredményez (pl. *kevésbé jó*), addig a negatív polaritású szentimentkifejezések mellett nem idéz elő feltétlenül polaritásváltást (pl. *kevésbé rossz*).

A csökkentő értelmű intenzifikáló elemeknél mutatkozó tendencia a negáló és az irreáló elemek esetében éppúgy megfigyelhető. Ez utóbbi elemek ugyanis szintén gyakoribb előfordulásúak a korpuszban a negatív polaritású fragmentumokban (pl. *nem jó*; *jó volna*).

A kapott eredmények arra mutatnak, hogy a magyar nyelvű szövegek automatikus szentimentelemzése során különös figyelmet kell majd fordítanunk a pozitív szentimentkifejezések mellett megjelenő negáló, irreáló, valamint csökkentő értelmű intenzifikáló elemekre, hiszen azok gyakorta okoznak az esetükben polaritásváltást.

A szentimentértéket befolyásoló egyik leggyakoribb, ugyanakkor az egyik nehezebben kezelhető nyelvi jelenség a negáció. Automatikus feldolgozását két alapvető sajátság is nehezíti: Egyrészt, a negáló elemekkel alkotott szerkezetek gyakorta nem

kompozicionálisak [4, 5]. Bár gyakran, de nem minden esetben okoznak polaritásváltást, csupán polaritásváltozást. A polaritásváltás az a jelenség, amelyben a teljes, értékelést megfogalmazó fragmentum polaritása és a bennfoglalt puszta szentimentkifejezés lexikai szintű polaritása egymással ellentétes. Ezzel szemben polaritásváltozásnak nevezzük azt az esetet, amikor a fragmentum polaritása csupán bizonyos mértékű elmozdulást mutat a szentimentkifejezés lexikai szintű polaritásához képest, vagy neutralizálódik [6, 7]. A negáció másik sajátága, amely megnehezíti annak kezelését, az az, hogy a negáló elem funkcióját a szövegekben számos kifejezés betöltheti (erről a későbbiekben részletesebben is lesz szó, l. lentebb).

Annak céljából, hogy pontosabb képet kaphassunk a negáló elemek szerepéről, megvizsgáltuk, hogy hány esetben okozzák ténylegesen a polaritás ellentétére változását. Megvizsgáltuk tehát minden negatív és pozitív szentimentkifejezés fragmentumbeli környezetét, a következő eredménnyel: A korpuszban előforduló 7554 pozitív polaritású szentimentkifejezés mellett összesen 1728 esetben fordul elő negáló elem, amely az esetek túlnyomó többségében, összesen 1554 kifejezésnél a polaritás megváltozását okozta. Hasonló eredményt kaptunk a negatív szentimentkifejezések vizsgálatával is. A negatív polaritású kifejezések közül összesen 1512 mellett jelenik meg a negáló elem, amely 1287 esetben a polaritás megváltozását idézte elő. Mindez arra mutat, hogy bár elméletileg a negáló elemek vagy teljes polaritásváltást, vagy csupán polaritásbeli elmozdulást idéznek elő, a gyakorlatban – legalábbis a korpuszunk adatai alapján – az előbbit tekinthetjük tendenciának.

A szófaji megoszlási arányok tekintetében különösen figyelemre méltónak találtuk a negáló elemek sajátosságait. Az annotált korpusz szófaji szempontú vizsgálata előtt azt feltételeztük, hogy a negáló elemként alapvetően tagadószókat (pl. *ne, sem, dehogy*) és a létige tagadó alakjait (*nincs, nincsen, sincs, sincsen*), emellett kisebb számban tagadó névutókat (pl. *hiányában, nélkül*), valamint néhány egyéb módosítószót (pl. *aligha, látszatra*) fogunk annotálni [8, 9]. A korpuszvizsgálat eredményei azonban arra mutatnak, hogy a negáló elemek nagyobb változatossággal szerepelnek kezdeti hipotézisünkénél. A korpuszban annotált összesen 3516 negációs token legnagyobb részét határozószók (2587), igék (468), névmások (145), valamint mellérendelő kötőszók (93) adták.

Annak céljából, hogy érzékeltessük a negáló elemek változatosságát, a negátorként annotált nyelvi elemek közül (3) alatt közlünk néhány példát.

- (3) hiánya, elillant, nélkülözi, bizarr lenne azt állítani, helyett, hiányoltam, semmi köze sincs, nulla, nem érezni benne, lespórolták

Megvizsgálva a csökkentő és a növelő értelmű intenzifikáló elemek szófaji megoszlását megállapítottuk, hogy amíg az előbbi funkcióját a legtöbbször melléknév tölti be mind a pozitív (31,44%), mind a negatív fragmentumokban (52,08%), addig az utóbbi funkciójában az esetek többségében (52,32% és 47,69%) határozószót találunk. Ugyanakkor, amint arra a fentebbi megoszlási arányok rámutatnak, a csökkentő értelmű intenzifikáló elemek esetében a szófaji megoszlási arányok nem azonosak a pozitív és a negatív polaritású fragmentumok között. A pozitív fragmentumokban ugyanis a különböző szófajok előfordulási arányai kiegyenlítettebbek a negatív fragmentu-

mokhoz képest, és bennük például a határozószók és a főnevek is jelentősebb szerephez jutnak a vizsgált funkcióban.

Az irreálók részletesebb vizsgálata során kiderült, hogy körülbelül kétharmaduk ige vagy határozószó. A határozószók esetében elsődlegesen a szójelentés hordozza az irreáló tartalmat (pl. *állítólag, valószínűleg, talán*), melyek párhuzamot mutatnak az úgynevezett episztemikus bizonytalanságot jelölő lexikai elemekkel [5]. Az igeéknél azonban kettősséget figyelhetünk meg: egyrészt magának az igeének a jelentése hordozza a bizonytalanságot (pl. *tűnik, hasonlít, imitál*) vagy szubjektivitást (pl. *érez, gondol*), másrészt morfológiai eszközök segítségével érhető el az irreáló tartalom. Ilyen például a lehetőséget kifejező -hAt toldalék, mely 115 esetben (11%) fordult elő a korpusz irreáló elemei között, és a feltételes mód, mely 291-szer (28%) jelent meg. A feltételes módhoz gyakran kapcsolódó *ha* és *mintha* kötőszavak is gyakori irreáló elemek a korpuszban, összesen 228 esetben (22%) találkozhattunk velük. A korpuszunkban megjelenő, nem a valóságot leíró nyelvi elemek szoros átfedést mutatnak az [5]-ben felvázolt, nyelvi bizonytalanságot jelölő elemekkel, így ezek összevető elemzésére a későbbiekben mindenképpen szeretnénk sort keríteni.

3.2. Felhasználási eredmények

Az annotációra támaszkodva bizonyos elemcsoportokból lexikonokat generáltunk, amelyek reményeink szerint javítani tudnak majd az automatikus szentimentelemzés eredményességén. A munka keretében pozitív és negatív szentimentlexikont, entitás- és aspektusszótárat, továbbá a különböző szentimentshifterek szótárait generáltuk, melyek alapvető statisztikai adatait az alábbi táblázat mutatja be:

2. táblázat. A korpuszból generált lexikonok alapvető statisztikai adatai

| lexikon | elemszám |
|----------------------------------|----------|
| Pozitív szentimentkifejezések | 2568 |
| Negatív szentimentkifejezések | 3343 |
| Entitások szótára | 2773 |
| Aspektusok szótára | 4781 |
| Negáló elemek | 95 |
| Növelő értelmű intenzifikálók | 744 |
| Csökkentő értelmű intenzifikálók | 199 |
| Irreálók | 195 |

A szentimentkifejezésekből generált lexikonok eredményességét szeretnénk összevetni a korábban, nem korpuszalapon készített szótáraink hatékonyságával is a jövőben.

Az entitások és az aspektusok szótárait a második annotálási szakaszhoz kidolgozott target-annotációnak köszönhetően volt lehetőségünk generálni, hiszen – ahogyan azt már korábban ismertettük (l. 2.2) – a korpuszban előforduló entitásokat, valamint azok különböző aspektusait eltérő annotációs taggel (target 1-20) láttuk el. Úgy véljük, ezek a szótárak különleges segítséget nyújthatnak majd hasonló, termékekkel kapcsolatban megfogalmazott véleményesszövegek automatikus szentimentelemzése során.

4. Az eredmények felhasználási lehetőségei

A dolgozatban a magyar nyelvű kézzel annotált szentimentkorpuszunk annotálásának tapasztalatait, valamint hasznosításának első eredményeit mutattuk be.

A korpusz második annotálási fázisában átdolgoztuk az annotálási elveket, majd a korpusz teljes szöveganyagát feldolgoztuk kézzel. A dolgozatban bemutattuk mindazokat a módosításokat, amelyeket a korpusz második annotálási fázisában alkalmaztunk az első annotálási fázis megoldásaihoz képest, a későbbi hasznosíthatóság javítása érdekében. Ezt követően részletesen tárgyaltuk a korpusz első vizsgálati és felhasználási eredményeit. A vizsgálat keretében elsősorban a különböző annotált elemek fragmentum-beli, valamint szófajok közötti megoszlását tártuk fel. Emellett az annotációra támaszkodva bizonyos elemcsoportokból lexikonokat generáltunk, amelyek reményeink szerint támogatni tudják majd az automatikus feldolgozói munka eredményességét. A dolgozatban a generált lexikonok alapvető statisztikai adatairól is számot adtunk.

A korpusz hozadékaira támaszkodva terveink között szerepel egy olyan automatikus elemző rendszer létrehozása, amely képes a szentimentkifejezéseket azok targetjeivel és shiftereivel összefüggésben hatékonyan kezelni.

Hivatkozások

1. Szabó M.K., Vincze V.: Egy magyar nyelvű szentimentkorpusz létrehozásának tapasztalatai. In: Tanács A., Varga V., Vincze V. (szerk.): XI. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2015), Szeged, Szegedi Tudományegyetem (2015) 219–226
2. Liu, B.: Sentiment Analysis and Opinion Mining. Draft (2012) Elérhető: <http://www.cs.uic.edu/~liub/FBS/SentimentAnalysis-and-OpinionMining.pdf>
3. Boucher, J., Osgood, C.: The Pollyanna hypothesis. *Journal of Verbal and Learning Behavior* 8(1) (1969) 1–8
4. Israel, M.: The pragmatics of polarity. In: Horn, L., Ward, G. (szerk.): *The Handbook of Pragmatics*, Oxford, Blackwell (2004) 701–723
5. Feldman, R., Rozenfeld, B., Breakstone, M.Y. SSA: A Hybrid Approach to Sentiment Analysis of Stocks (2010) Elérhető: http://web.mit.edu/michab/www/ISCOL_Paper.pdf

6. Szabó M. K.: A polaritásváltás és -változás kezelési lehetőségei a szentimentelemzésben. In: Tavaszi Szél Konferencia, Eger (2015a) Megjelenés előtt
7. Szabó M. K.: A nyelvi értékelés mibenléte a számítógépes értékeléselemzés (szentimentelemzés) szempontjából. In: Lingdok – Nyelvész-doktoranduszok dolgozatai, Szeged (2015b) Megjelenés előtt
8. Szabó M. K., Vincze V.: A negáló szentimentshifterek kezelési kérdései a magyar nyelvű szövegek szentimentelemzésében. In: XXV. Magyar Alkalmazott Nyelvészeti Kongresszus előadásai (2015) Publikálásra benyújtva
9. Pete, I.: Az állító és tagadó mondatok szinonimiája a magyarban. Magyar Nyelv 95/3. (1999) 305–312
10. Vincze, V.: Uncertainty Detection in Hungarian Texts. In: Proceedings of COLING 2014, Dublin (2014) 1844–1853

Forrás

<http://divany.hu/>